cen
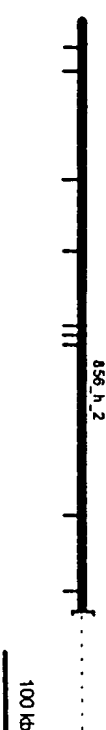
D12S1328

P46F2t7

P22D22sp6

B157K14sp6

P168L1T7

B139I11sp6

D12S1332

168L1sp6

B96I23sp6

B147O14CAG

B88A9t7
B42P18sp6

B42B1t7

B147O14sp6

P287C7sp6
B162N10t7
P162P23t7
B42B1sp6

101D15t7

ALDex6/12
102D18t7

D12S1333

B255C05t7

B255C05sp6
B167N19sp6
B087C05t7

B167N19t7

B175g1t7

B110L21sp6
B78E14t7

P46F2

P22D22

B157K14

B139I11

P168L1

P168A22

B96I23

Y636_E_7

P88A9

P162N10

B147O14

P287C7

P42P18

B42B1

P65Z2

B162P23

P273B7

P101D15

P102D18

P14B2

B255C05

P282O14

B167N19

B087C05

B175G1

B110L21

B78E14

Y884_h_11

792_b_4

Y731_b_7

Y919_a_2

856_h_2

100 kb

FIGURE 1

t7
sp6

1 TTGGTAGCAACGGAAACCGCCGGCGGCGCGTTTCGGCCCGGCTCCCGGCCGGCTCCTTGGTC

61 TCGGCGGGCCTCCCCGCCCCTTCGTCGTCGTCCTTCTCCCCCTCCCCAGCCCCGGGCGCCC

121 CTCCGGCCGCGCCAACCCCGCCGCCTCCCCGCTCCGCGCCCCGTGCGTCCCCGCCGCGTTCCG

181 GCGTCTCCTTGGCGCGCCCGGCTCCCGGCTGTCCCCGCCCGGCGTGCGAGCCGGTGTATG

         SCA2-A   →
241 GGCCCCTCACCATGTCGCTGAAGCCCCAGCAGCAGCAGCAGCAGCAGCAGCAACAGCAGC

301 AGCAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCCGCCGCCCGCGGCTGCCAATGTCCGCA
                                             ← SCA2-B

361 AGCCCGGCGGCAGCGGCCCTTCTAGCGTCGCCCGCCGCCCCGCGCCCTTCGCCGTCCTCGTCCT

421 CGGTCTCCTCGTCCTCGGCCACGGCTCCCTCCTCGGTGGTCGCGGCGACCTCCGGCGGCG

          |
          ↓
481 GGAGGCCCGGCCTGGGCAGGTGGGTGTCGGCACCCC

FIGURE 2

FIGURE 3

AGE OF ONSET

NUMBER OF (CAG) REPEATS

FIGURE 4

F1.1-F1.7 5' ▬▬▬ AAA 3'
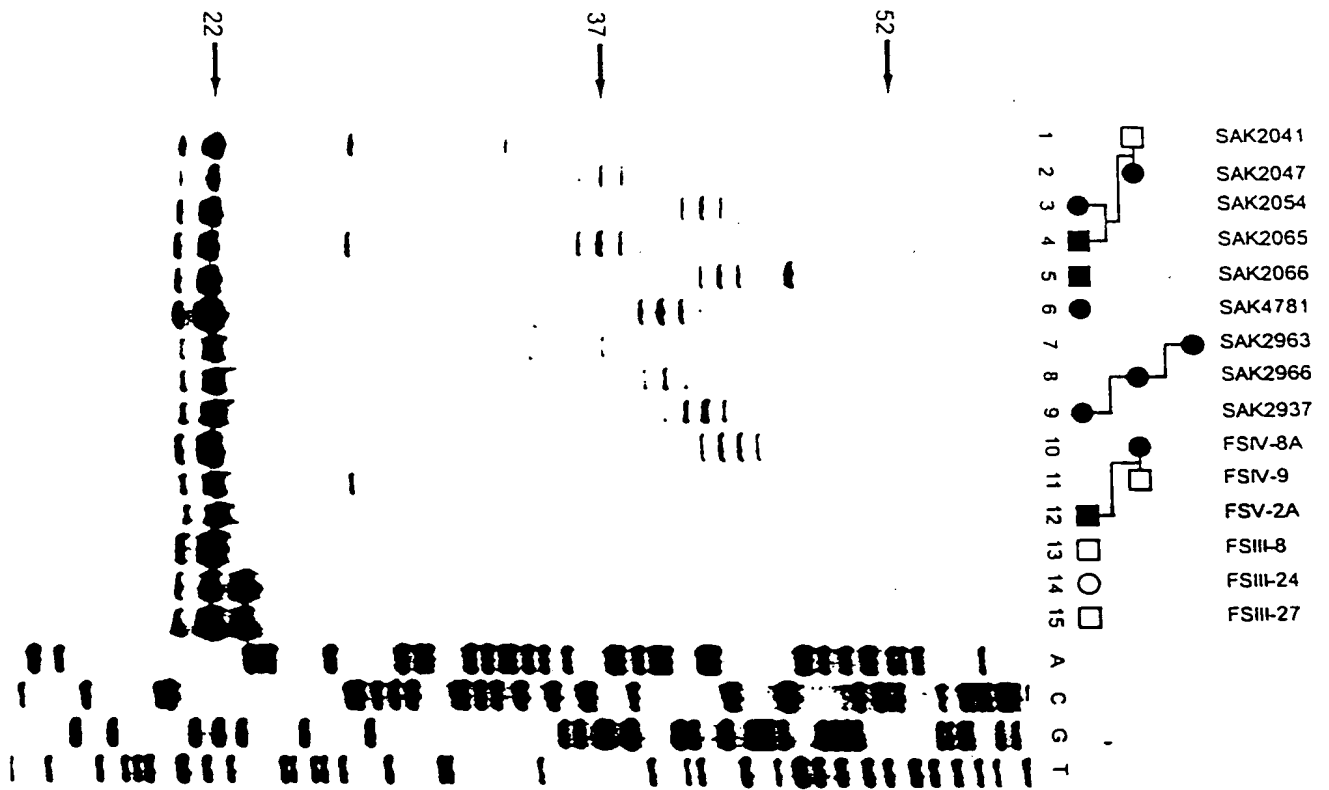
F2.1-F2.3 5' ▬▬▬ AAA 3'

S1 5' ▬▬▬ 3'

S2 5' ▬ 3'

300 bp
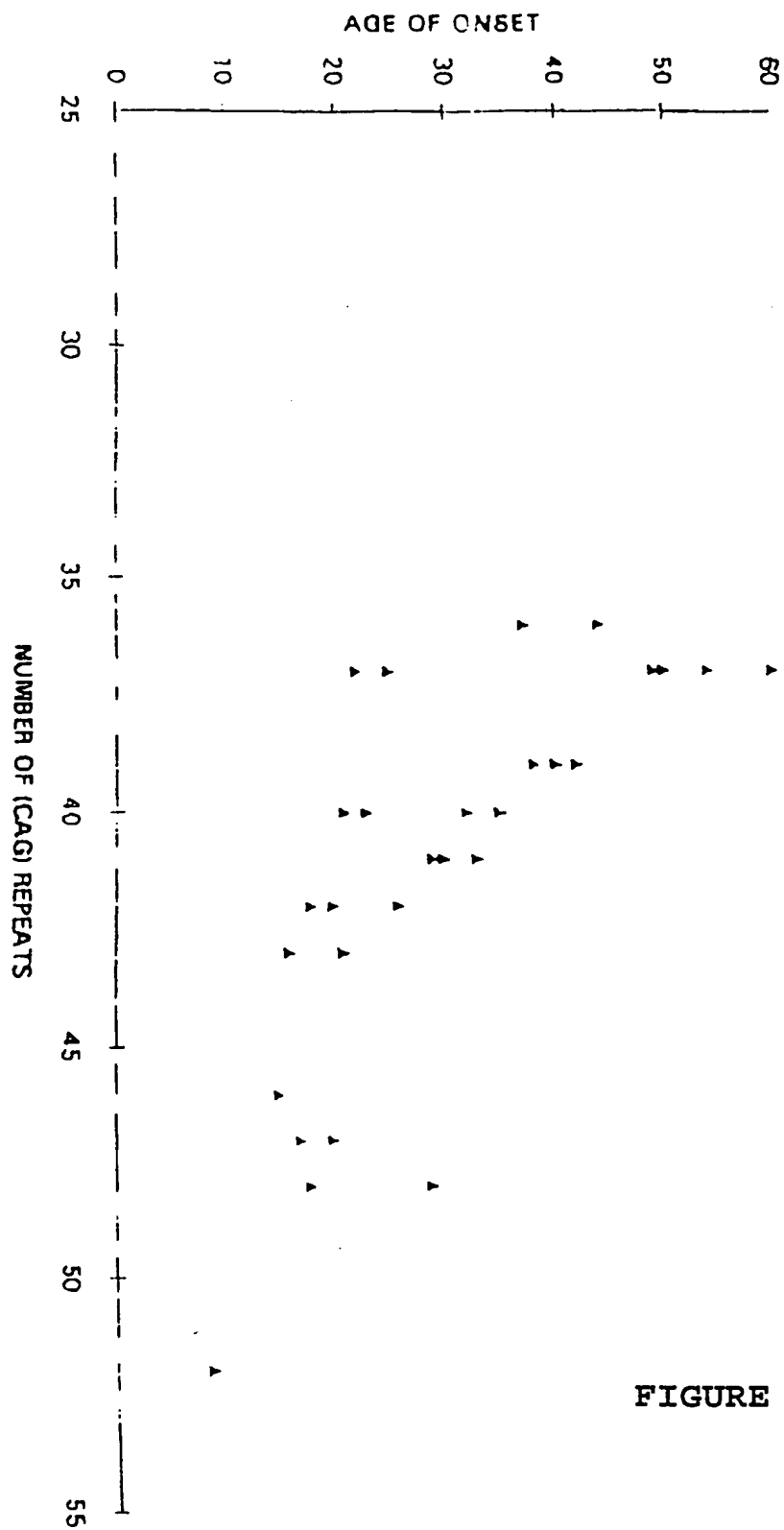
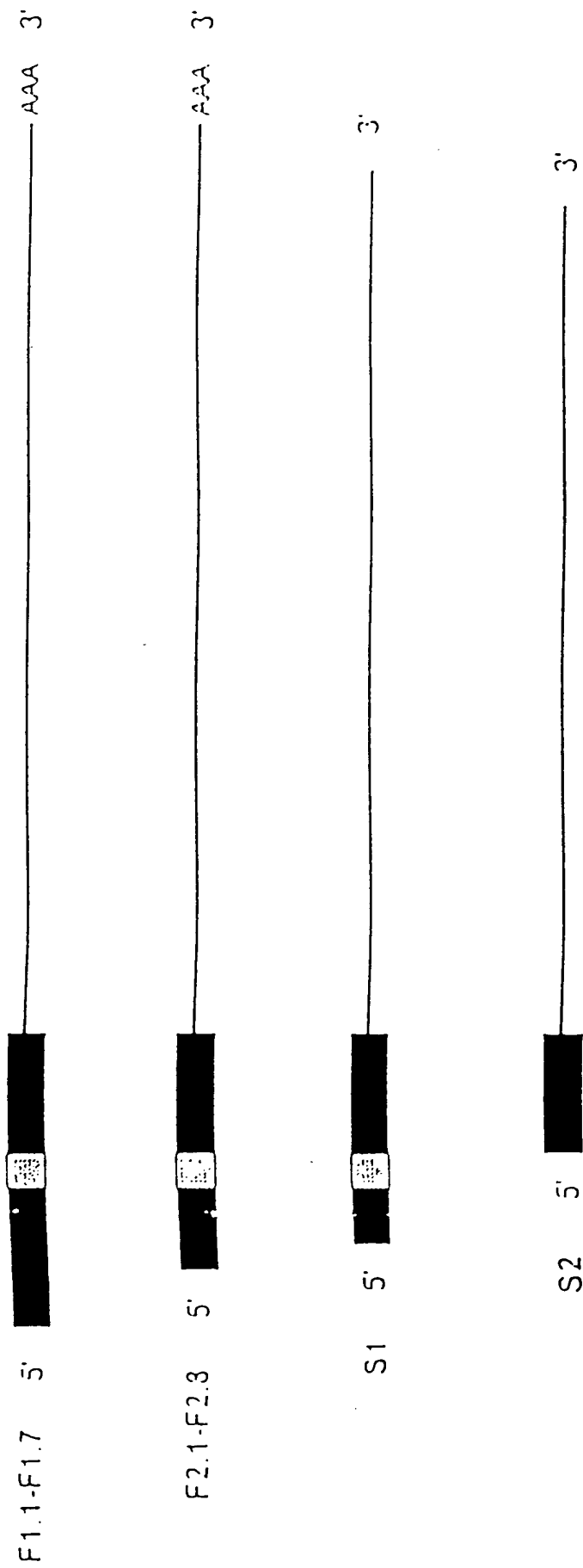FIGURE 5

```
   1  ACCCCCGAGAAAGCAACCCAGCGCGCCGCCCGCTCCTCACGTGTCCCTCCCGGCCCCGGG      60
  61  GCCACCTCACGTTCTGCTTCCGTCTGACCCCTCCGACTTCCGGTAAAGAGTCCCTATCCG     120

 121  CACCTCCGCTCCCACCCGGCGCCTCGGCGCGCCCGCCCTCCGATGCGCTCAGCGGCCGCA     180
   1                                              M  R  S  A  A  A      6
 181  GCTCCTCGGAGTCCCGCGGTGGCCACCGAGTCTCGCCGCTTCGCCGCAGCCAGGTGGCCC     210
   7   A  P  R  S  P  A  V  A  T  E  S  R  R  F  A  A  A  R  W  P     26
 241  GGGTGGCGCTCGCTCCAGCGGCCGGCGCGGCGGAGCGGGCGGGGCGGCGGTGGCGCGGCC     300
  27   G  W  R  S  L  Q  R  P  A  R  R  S  G  R  G  G  G  A  A       46
 301  CCGGGACCGTATCCCTCCGCCGCCCCTCCCCCGCCCGGCCCCGGCCCCCCTCCCTCCCGG     360
  47   P  G  P  Y  P  S  A  A  P  P  P  P  G  P  G  P  P  S  R       66
 361  CAGAGCTCGCCTCCCTCCGCCTCAGACTGTTTTGGTAGCAACGGCAACGGCGGCGGCGCG     420
  67   Q  S  S  P  P  S  A  S  D  C  F  G  S  N  G  N  G  G  A       86
 421  TTTCGGCCCGGCTCCCGGCGGCTCCTTGGTCTCGGCGGGCCTCCCCGCCCCTTCGTCGTC     480
  87   F  R  P  G  S  R  R  L  L  G  L  G  G  P  P  R  P  F  V  V    106
 481  GTCCTTCTCCCCCTCGCCAGCCCGGGCGCCCCCTCCGGCCGCGCCAACCCGCGCCTCCCCG     540
 107   V  L  L  P  L  A  S  P  G  A  P  P  A  A  P  T  R  A  S  P    126
 541  CTCGGCGCCCGTGCGTCCCCGCCGCGTTCCGGCGTCTCCTTGGCGCGCCCGGCTCCCGGC     600
 127   L  G  A  R  A  S  P  P  R  S  G  V  S  L  A  R  P  A  P  G    146
```

                                          ───────SCA2-A────────►
```
 601  TGTCCCGCCCGGCGTGCGAGCCGGTGTATGGGCCCCTCACCATGTCGCTGAAGCCCCAG     660
 147   C  P  A  C  E  P  V  Y  G  P  L  T  M  S  L  K  P  Q         166
 661  CAGCAGCAGCAGCAGCAGCAGCAACAGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCAG     720
 167   Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q   186
```

                       ◄─────SCA2-B──────
```
 721  CAGCCGCCGCCCGCGGCTGCCAATGTCCGCAAGCCCGGCGGCAGCGGCCTTCTAGCGTCG     780
 187   Q  P  P  P  A  A  A  N  V  R  K  P  G  G  S  G  L  L  A  S   206
 781  CCCGCCGCCGCGCCTTCGCCGTCCTCGTCCTCGGTCTCCTCGTCCTCGGCCACGGCTCCC     840
 207   P  A  A  A  P  S  P  S  S  S  V  S  S  S  S  A │ T  A  P     226
 841  TCCTCGGTGGTCGCGGCGACCTCCGGCGGCGGGAGGCCCGGCCTGGGCAGAGGTCGAAAC     900
 227   S  S  V  V  A  A  T  S  G  G  G  R  P  G  L  G  R  G  R  N   246
 901  AGTAACAAAGGACTGCCTCAGTCTACGATTTCTTTTGATGGAATCTATGCAAATATGAGG     960
 247   S  N  K  G  L  P  Q  S  T  I  S  F  D  G  I  Y  A  N  M  R   266
 961  ATGGTTCATATACTTACATCAGTTGTTGGCTCCAAATGTGAAGTACAAGTGAAAAATGGA    1020
 267   M  V  H  I  L  T  S  V  V  G  S  K  C  E  V  Q  V  K  N  G   286
```

                                      ────────SCA2-14B────────
```
1021  GGTATATATGAAGGAGTTTTTTAAAACTTACAGTCCGAAGTGTGATTTGGTACTTGATGCC    1080
 287   G  I  Y  E  G  V  F  K  T  Y  S  P  K  C  D  L  V  L  D  A   306
1081  GCACATGAGAAAGTACAGAATCCAGTTCGGGGCCGAAACGTGAAGAAATAATGGAGAGT    1140
 307   A  H  E  K  S  T  E  S  S  S  G  P  K  R  E  E  I  M  E  S   326
1141  ATTTTGTTCAAATGTTCAGACTTTGTTGTGGTACAGTTTAAAGATATGGACTCCAGTTAT    1200
 327   I  L  F  K  C  S  D  F  V  V  V  Q  F  K  D  M  D  S  S  Y   346
1201  GCAAAAAGAGATGCTTTTACTGACTCTGCTATCAGTGCTAAAGTGAATGGCGAACACAAA    1260
 347   A  K  R  D  A  F  T  D  S  A  I  S  A  K  V  N  G  E  H  K   366
1261  GAGAAGGACCTGGAGCCCTGGATGCAGGTGAACTCACAGCCAATGAGGAACTTGAGGCT    1320
 367   E  K  D  L  E  P  W  D  A  G  E  L  T  A  N  E  E  L  E  A   386
1321  TTGGAAAATGACGTATCTAATGGATGGGATCCCAATGATATGTTTCGATATAATGAAGAA    1380
 387   L  E  N  D  V  S  N  G  W  D  P  N  D  M  F  R  Y  N  E  E   406
1381  AATTATGGTGTAGTGTCTACGTATGATAGCAGTTTATCTTCGTATACAGTGCCCTTAGAA    1440
 407   N  Y  G  V  V  S  T  Y  D  S  S  L  S  S  Y  T  V  P  L  E   426
1441  AGAGATAACTCAGAAGAATTTTTTAAAACGGGAAGCAAGGGCAAACCAGTTAGCAGAAGAA    1500
 427   R  D  N  S  E  E  F  L  K  R  E  A  R  A  N  Q  L  A  E  E   446
```

FIGURE 6A

```
1501 ATTGAGTCAAGTGCCCAGTACAAAGCTCGAGTGGCCCTGGAAAATGATGATAGGAGTGAG 1560
 447  I  E  S  S  A  Q  Y  K  A  R  V  A  L  E  N  D  D  R  S  E  466
1561 GAAGAAAAATACACAGCAGTTCAGAGAAATTCCAGTGAACGTGAGGGGCACAGCATAAAC 1620
 467  E  E  K  Y  T  A  V  Q  R  N  S  S  E  R  E  G  H  S  I  N  486
1621 ACTAGGGAAAATAAATATATTCCTCCTGGACAAAGAAATAGAGAAGTCATATCCTGGGGA 1680
 487  T  R  E  N  K  Y  I  P  P  G  Q  R  N  R  E  V  I  S  W  G  506
1681 AGTGGGAGACAGAATTCACCGCGTATGGGCCAGCCTGGATCGGGCTCCATGCCATCAAGA 1710
 507  S  G  R  Q  N  S  P  R  M  G  Q  P  G  S  G  S  M  P  S  R  526
1741 TCCACTTCTCACACTTCAGATTTCAACCCGAATTCTGGTTCAGACCAAAGAGTAGTTAAT 1800
 527  S  T  S  H  T  S  D  F  N  P  N  S  G  S  D  Q  R  V  V  N  546
1801 GGAGGTGTTCCCTGGCCATCGCCTTGCCCATCTCCTTCCTCTCGCCCACCTTCTCGCTAC 1860
 547  G  G  V  P  W  P  S  P  C  P  S  P  S  S  R  P  P  S  R  Y  566
1861 CAGTCAGGTCCCAACTCTCTTCCACCTCGGGCAGCCACCCCTACACGGCCGCCCTCCAGG 1920
 567  Q  S  G  P  N  S  L  P  P  R  A  A  T  P  T  R  P  P  S  R  586
1921 CCCCCCTCGCGGGCCATCCAGACCCCCGTCTCACCCCTCTGCTCATGGTTCTCCAGCTCCT 1980
 587  P  P  S  R  P  S  R  P  P  S  H  F  S  A  H  G  S  P  A  P  606
1981 GTCTCTACTATGCCTAAACGCATGTCTTCAGAAGGGCCTCCAAGGATGTCCCCAAAGGCC 2040
 607  V  S  T  M  P  P  R  M  S  S  E  G  P  P  R  M  S  P  K  A  626
2041 CAGCGACATCCTCGAAATCACAGAGTTTCTGCTGGGAGGGGTTCCATATCCAGTGGCCTA 2100
 627  Q  R  H  P  R  N  H  R  V  S  A  G  R  G  S  I  S  S  G  L  646
2101 GAATTTGTATCCCACAACCCACCCAGTGAAGCAGCTACTCCTCCAGTAGCAAGGACCAGT 2160
 647  E  F  V  S  H  N  P  P  S  E  A  A  T  P  P  V  A  R  T  S  666
2161 CCCTCGGGGGGAACGTGGTCATCAGTGGTCAGTGGGGTTCCAAGATTATCCCCTAAAACT 2220
 667  P  S  G  G  T  W  S  S  V  V  S  G  V  P  R  L  S  P  K  T  686
2221 CATAGACCCAGGTCTCCCAGACAGAACAGTATTGGAAATACCCCCAGTGGGCCAGTTCTT 2280
 687  H  R  P  R  S  P  R  Q  N  S  I  G  N  T  P  S  G  P  V  L  706
2281 GCTTCTCCCCAAGCTGGTATTATTCCAACTGAAGCTGTTGCCATGCCTATTCCAGCTGCA 2340
 707  A  S  P  Q  A  G  I  I  P  T  E  A  V  A  M  P  I  P  A  A  726
2341 TCTCCTACGCCTGCTAGTCCTGCATCGAACAGAGCTGTTACCCCTTCTAGTGAGGCTAAA 2400
 727  S  P  T  P  A  S  P  A  S  N  R  A  V  T  P  S  S  E  A  K  746
2401 GATTCCAGGCTTCAAGATCAGAGGCAGAACTCTCCTGCAGGGAATAAAGAAAATATTAAA 2460
 747  D  S  R  L  Q  D  Q  R  Q  N  S  P  A  G  N  K  E  N  I  K  766
2461 CCCAATGAAACATCACCTAGCTTCTCAAAGCTGAAAACAAAGGTATATCACCAGTTGTT 2520
 767  P  N  E  T  S  P  S  F  S  K  A  E  N  K  G  I  S  P  V  V  786
2521 TCTGAACATAGAAAACAGATTGATGATTTAAAGAAATTTAAGAATGATTTTAGGTTACAG 2580
 787  S  E  H  R  K  Q  I  D  D  L  K  K  F  K  N  D  F  R  L  Q  806
2581 CCAAGTTCTACTTCTGAATCTATGGATCAACTACTAAACAAAAATAGAGAGGGAGAAAAA 2640
 807  P  S  S  T  S  E  S  M  D  Q  L  L  N  K  N  R  E  G  E  K  826
2641 TCAAGAGATTTGATCAAAGACAAAATTGAACCAAGTGCTAAGGATTCTTTCATTGAAAAT 2700
 827  S  R  D  L  I  K  D  K  I  E  P  S  A  K  D  S  F  I  E  N  846
2701 AGCAGCAGCAACTGTACCAGTGGCAGCAGCAAGCCGAATAGCCCCAGCATTTCCCCTTCA 2760
 847  S  S  S  N  C  T  S  G  S  S  K  P  N  S  P  S  I  S  P  S  866
2761 ATACTTAGTAACACGGAGCACAAGAGGGGACCTGAGGTCACTTCCCAAGGGGTTCAGACT 2820
 867  I  L  S  N  T  E  H  K  R  G  P  E  V  T  S  Q  G  V  Q  T  886
2821 TCCAGCCCAGCATGTAAACAAGAGAAAGACGATAAGGAAGAGAAGAAAGACGCAGCTGAG 2880
 887  S  S  P  A  C  K  Q  E  K  D  D  K  E  E  K  K  D  A  A  E  906
2881 CAAGTTAGGAAATCAACATTGAATCCCAATGCAAAGGAGTTCAACCCACGTTCCTTCTCT 2940
 907  Q  V  R  K  S  T  L  N  P  N  A  K  E  F  N  P  R  S  F  S  926
2941 CAGCCAAAGCCTTCTACTACCCCAACTTCACCTCGGCCTCAAGCACAACCTAGCCCATCT 3000
 927  Q  P  K  P  S  T  I  P  T  S  P  R  P  Q  A  Q  P  S  P  S  946
3001 ATGGTGGGTCATCAACAGCCAACTCCAGTTTATACTCAGCCTGTTTGTTTTGCACCAAAT 3060
 947  M  V  G  H  Q  Q  P  T  P  V  Y  T  Q  P  V  C  F  A  P  N  966
3061 ATGATGTATCCAGTCCCAGTGAGCCCAGGCGTGCAACCTTTATACCCAATACCTATGACG 3120
 967  M  M  Y  P  V  P  V  S  P  G  V  Q  P  L  Y  P  I  P  M  T  986
```

FIGURE 6B

```
3121  CCCATGCCAGTGAATCAAGCCAAGACATATAGAGCAGTACCAAATATGCCCCAACAGCGG  3180
 987   P  M  P  V  N  Q  A  K  T  Y  R  A  V  P  N  M  P  Q  Q  R  1006
3181  CAAGACCAGCATCATCAGAGTGCCATGATGCACCCAGCGTCAGCAGCGGGCCCACCGATT  3240
1007   Q  D  Q  H  H  Q  S  A  M  M  H  P  A  S  A  A  G  P  P  I  1026
3241  GCAGCCACCCCCACCAGCTTACTCCACGCAATATGTTGCCTACAGTCCTCAGCAGTTCCCA  3300
1027   A  A  T  P  P  A  Y  S  T  Q  Y  V  A  Y  S  P  Q  Q  F  P  1046
3301  AATCAGCCCCTTGTTCAGCATGTGCCACATTATCAGTCTCAGCATCCTCATGTCTATAGT  3360
1047   N  Q  P  L  V  Q  H  V  P  H  Y  Q  S  Q  H  P  H  V  Y  S  1066
3361  CCTGTAATACAGGGTAATGCTAGAATGATGGCACCACCAACACACGCCCAGCCTGGTTTA  3420
1067   P  V  I  Q  G  N  A  R  M  M  A  P  P  T  H  A  Q  P  G  L  1086
3421  GTATCTTCTTCAGCAACTCAGTACGGGGGCTCATGAGCAGACGCATGCGATGTATGCATGT  3480
1087   V  S  S  S  A  T  Q  Y  G  A  H  E  Q  T  H  A  M  Y  A  C  1106
3481  CCCAAATTACCATACAACAAGGAGACAAGCCCTTCTTTCTACTTTGCCATTTCCACGGGC  3540
1107   P  K  L  P  Y  N  K  E  T  S  P  S  F  Y  F  A  I  S  T  G  1126
3541  TCCCTTGCTCAGCAGTATGCGCACCCTAACGCTACCCTGCACCCACATACTCCACACCCT  3600
1127   S  L  A  Q  Q  Y  A  H  P  N  A  T  L  H  P  H  T  P  H  P  1146
3601  CAGCCTTCAGCTACCCCCACTGGACAGCAGCAAAGCCAACATGGTGGAAGTCATCCTGCA  3660
1147   Q  P  S  A  T  P  T  G  Q  Q  Q  S  Q  H  G  G  S  H  P  A  1166
3661  CCCAGTCCTGTTCAGCACCATCAGCACCAGGCCGCCCAGGCTCTCCATCTGGCCAGTCCA  3720
1167   P  S  P  V  Q  H  H  Q  H  Q  A  A  Q  A  L  H  L  A  S  P  1186
3721  CAGCAGCAGTCAGCCATTTACCACGCGGGGCTTGCGCCAACTCCACCCTCCATGACACCT  3780
1187   Q  Q  Q  S  A  I  Y  H  A  G  L  A  P  T  P  P  S  M  T  P  1206
3781  GCCTCCAACACGCAGTCGCCACAGAATAGTTTCCCAGCAGCACAACAGACTGTCTTTACG  3840
1207   A  S  N  T  Q  S  P  Q  N  S  F  P  A  A  Q  Q  T  V  F  T  1226
3841  ATCCATCCTTCTCACGTTCAGCCGGCGTATACCAACCCCACCCCACATGGCCCACGTACCT  3900
1227   I  H  P  S  H  V  Q  P  A  Y  T  N  P  P  H  M  A  H  V  P  1246
3901  CAGGCTCATGTACAGTCAGGAATGGTTCCTTCTCATCCAACTGCCCATGCGCCAATGATG  3960
1247   Q  A  H  V  Q  S  G  M  V  P  S  H  P  T  A  H  A  P  M  M  1266
3961  CTAATGACGACACAGCCACCCGGCGGTCCCCAGGCCGCCCTCGCTCAAAGTGCACTACAG  4020
1267   L  M  T  T  Q  P  P  G  G  P  Q  A  A  L  A  Q  S  A  L  Q  1286
4021  CCCATTCCAGTCTCGACAACAGCGCATTTCCCCTATATGACGCACCCTTCAGTACAAGCC  4080
1287   P  I  P  V  S  T  T  A  H  F  P  Y  M  T  H  P  S  V  Q  A  1306
4081  CACCACCAACAGCAGTTGTAAGGCTGCCCTGGAGGAACCGAAAGGCCAAATTCCCTCCTC  4140
1307   H  H  Q  Q  L  .                                             1326
4141  CCTTCTACTGCTTCTACCAACTGGAAGCACAGAAAACTAGAATTTCATTTATTTTGTTTT  4200
4201  TAAAATATATATGTTGATTTCTTGTAACATCCAATAGGAATGCTAACAGTTCACTTGCAG  4260
4261  TGGAAGATACTTGGACCGAGTAGAGGCATTTAGGAACTTGGGGGCTATTCCATAATTCCA  4320
4321  TATGCTGTTTCAGAGTCCCGCAGGTACCCCAGCTCTGCTTGCCGAAACTGGAAGTTATTT  4380
4381  ATTTTTTAATAACCCTTGAAAGTCATGAACACATCAGCTAGCAAAGAAGTAACAAGAGT  4440
4441  GATTCTTGCTGCTATTACTGCTAAAAAAAAAAAAAAAAAAA  4481
```

FIGURE 6C

```
                    1                                                    50
Ataxin-2            VYGPLTMSLK PQQQQQQQQQ QQQQQQQQQQ QQQPPPAAAN VRKPGGSGLL
Mouse Ataxin-2      HEGPLTMSLK PQPQ...... .......... ....PPAPAT GRKPGG.GLL
A2RP               ........LA PQPPPPQQHQ ER........ .......... ..........
     Consensus     -------L- PQ------- ---------- ---------- ----------


                    51                                                   100
Ataxin-2            ASPAAAPSPS SSSVSSSSAT APSSVVA... ATSGGGRPGL GRGRNSNKGL
Mouse Ataxin-2      SSPGAAP.AS AAVTSASVVP APAAPVASSS AAAGGGRPGL GRGRNSSKGL
A2RP               ..PGAAAIGS A......... .......... .......... .RGQSTGKGP
     Consensus     --P-AA---S ---------- ---------- ---------- -RG----KG-


                    101                                                  150
Ataxin-2            PQSTISFDGI YANMRMVHIL TSVVGSKCEV QVKNGGIYEG VFKTYSPKCD
Mouse Ataxin-2      PQPTISFDGI YANVRMVHIL TSVVGSKCEV QVKNGGIYEG VFKTYSPKCD
A2RP               PQSPV.FEGV YNNSRMLHFL TAVVGSTCDV KVKNGTTYEG IFKTLSSKFE
     Consensus     PQ----F-G- Y-N-RM-H-L T-VVGS-C-V -VKNG--YEG -FKT-S-K--


                    151                                                  200
Ataxin-2            LVLDAAHEKS TESSSGPKRE EIMESILFKC SDFVVVQFKD MDSSYAKRDA
Mouse Ataxin-2      LVLDAAHEKS TESSSGPKRE EIMESVLFKC SDFVVVQFKD TDSSYARRDA
A2RP               LAVDAVHRKA SEPAGGPRRE DIVDTHVFKP SDVMLVHFRN VDFNYATKDK
     Consensus     L--DA-H-K- -E---GP-RE -I-----FK- SD---V-F-- -D--YA--D-


                    201                                                  250
Ataxin-2            FTDSAIS..A KVNGEHKEKD LEPWDAGELT ANEELEALEN DVSNGWDPND
Mouse Ataxin-2      FTDSALS..A KVNGEHKEKD LEPWDAGELT ASEELE.LEN DVSNGWDPND
A2RP               FTDSAIAHNS KVNGEHKEKV LQRWEGGD.S NSDDYD.LES DMSNGWDPNE
     Consensus     FTDSA----- KVNGEHKEK- L--W--G--- ------LE- D-SNGWDPN-


                    251                                                  300
Ataxin-2            MFRYNEEHYG VVSTYDSSLS SYTVPLERDN SEEFLKREAR ANQLAEEIES
Mouse Ataxin-2      MFRYNEEHYG VVSTYDSSLS SYTVPLERDN SEEFLKREAR ANQLAEEIES
A2RP               MFKFNEENYG VKTTYDSSLS SYTVPLEKDN SEEFRQRELR AAQLAREIES
     Consensus     MF--HEEHYG V--TYDSSLS SYTVPLE-DN SEEF--RE-R A-QLA-EIES


                    301                                                  350
Ataxin-2            SAQYKARVAL ENDD.RSEEE KYTAVQRNSS EREGHSINTR ENKYIPPGQR
Mouse Ataxin-2      SAQYKARVAL ENDD.RSEEE KYTAVQRNCS DREGHGPNTR DNKYIPPGQR
A2RP               SPQYRLRINM ENDDGRTEEE KHSAVQRQGS GRESPSLASR EGKYIP....
     Consensus     S-QY--R-A- ENDD-R-EEE K--AVQR--S -RE------R --KYIP----


                    351
Ataxin-2            HR
Mouse Ataxin-2      HR
A2RP               ..
     Consensus     --
```
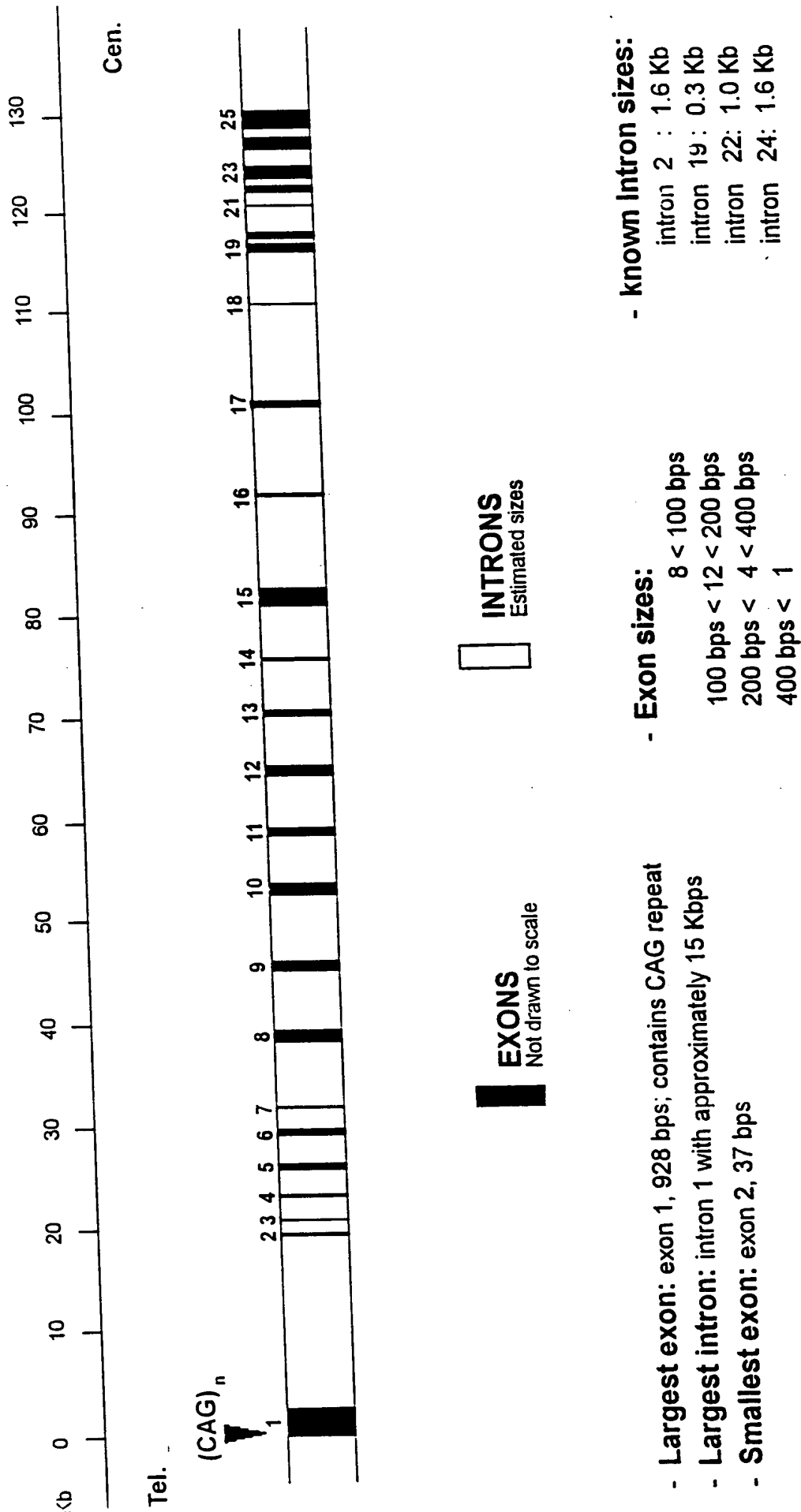
FIGURE 7

SCA2 Gene

FIGURE 8

EXONS
Not drawn to scale

INTRONS
Estimated sizes

- **Largest exon:** exon 1, 928 bps; contains CAG repeat
- **Largest intron:** intron 1 with approximately 15 Kbps
- **Smallest exon:** exon 2, 37 bps

- **Exon sizes:**

  8 < 100 bps
  100 bps < 12 < 200 bps
  200 bps < 4 < 400 bps
  400 bps < 1

- **known Intron sizes:**

  intron 2 : 1.6 Kb
  intron 19 : 0.3 Kb
  intron 22: 1.0 Kb
  intron 24: 1.6 Kb